Abstract

When one trained in speech perception hears the name Karl Kryter, they immediately think of the AI standard, which Kryter ferried through the standards process. Today the AI has been studied to the point that we are beginning to understand both its beauty, and its fundamental limitations. The AI was created by Fletcher and Galt, as reported first in JASA in 1950, and in Fletcher's two books. It is based on two important intertwined observations: 1) that the average consonant error Pe is exponential in an average of critical band signal to noise ratios (i.e., $P_e(SNR) = 0.015^{AI}$) and 2) the total error is the product of independent critical band errors. While AI theory has been verified many times over, the real question is "Why does it work?" In this talk I will answer this question, and in the process expose the AIs key weaknesses. AI theory has guided us to our modern understanding of speech perception in normal and hearing impaired ears. The publications supporting this work date from 1994-2014 and may be found at http://auditorymodels.org/ @ *Allen's Publications*.

Karl Kryter: The evolution of the Articulation Index

Past, Present & Future

Jont B Allen
UIUC-ECE Urbana IL, USA

Univ. of IL, Beckman Inst., Urbana IL

May 19, 2015

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs ...) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins $\Sigma 8$
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins $\Sigma 13$
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his "missing bands"
- Summary + Conclusions 2 mins $\Sigma 15$

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs ...) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins $\Sigma 8$
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins $\Sigma 13$
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his "missing bands"
- Summary + Conclusions 2 mins $\Sigma 15$

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins Σ4
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs . . .) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins Σ13
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his “missing bands”
- Summary + Conclusions 2 mins Σ15

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs ...) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins $\Sigma 8$
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins $\Sigma 13$
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his “missing bands”
- Summary + Conclusions 2 mins $\Sigma 15$

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins Σ4
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs . . .) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins Σ13
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his "missing bands"
- Summary + Conclusions 2 mins Σ15

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs ...) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins $\Sigma 8$
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins $\Sigma 13$
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his “missing bands”
- Summary + Conclusions 2 mins $\Sigma 15$

1. Objectives of the UIUC HSR Group

- Identify acoustic cues for >100 CV tokens
  - a) Onsets, b) Frequency edges, c) Duration, d) Voicing, ...
- Method: Measure CV confusions in ≈50 *normal hearing* ears
  - Show that consonant recognition is *binary*
  - Explain the *token SNR threshold*: SNR_0
  - Explain why $P_e(SNR{-}SNR_0) = \{0, 1\}$
- Explain how the AI works: *individual-differences*
  - *Hypothesis:* The AI's log-error vs. SNR follows from the distribution of binary thresholds
- Whats going on in the *Hearing Impared* ears?
  - *Hypothesis:* Its not about "listening in the gaps!"

1. Objectives of the UIUC HSR Group

- Identify acoustic cues for >100 CV tokens
  - a) Onsets, b) Frequency edges, c) Duration, d) Voicing, ...
- Method: Measure CV confusions in ≈50 *normal hearing* ears
  - Show that consonant recognition is *binary*
  - Explain the *token SNR threshold*: SNR_0
  - Explain why $P_e(SNR - SNR_0) = \{0, 1\}$
- Explain how the AI works: *individual-differences*
  - *Hypothesis:* The AI's log-error vs. SNR follows from the distribution of binary thresholds
- Whats going on in the *Hearing Impared* ears?
  - *Hypothesis:* Its not about "listening in the gaps!"

1. Objectives of the UIUC HSR Group

- Identify acoustic cues for >100 CV tokens
  - a) Onsets, b) Frequency edges, c) Duration, d) Voicing, ...
- Method: Measure CV confusions in ≈50 *normal hearing* ears
  - Show that consonant recognition is *binary*
  - Explain the *token SNR threshold*: SNR_0
  - Explain why $P_e(SNR - SNR_0) = \{0, 1\}$
- Explain how the AI works: *individual-differences*
  - *Hypothesis:* The AI's log-error vs. SNR follows from the distribution of binary thresholds
- Whats going on in the *Hearing Impared* ears?
  - *Hypothesis:* Its not about "listening in the gaps!"

1. Objectives of the UIUC HSR Group

- Identify acoustic cues for >100 CV tokens
  - a) Onsets, b) Frequency edges, c) Duration, d) Voicing, ...
- Method: Measure CV confusions in ≈50 *normal hearing* ears
  - Show that consonant recognition is *binary*
  - Explain the *token SNR threshold*: SNR_0
  - Explain why $P_e(SNR - SNR_0) = \{0, 1\}$
- Explain how the AI works: *individual-differences*
  - *Hypothesis:* The AI's log-error vs. SNR follows from the distribution of binary thresholds
- Whats going on in the *Hearing Impared* ears?
  - *Hypothesis:* Its not about "listening in the gaps!"

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs . . .) (>1950)
  - G.A. Miller 1951+

2. Historical Speech Studies

- Lord Rayleigh 1908 and George Campbell 1910
  - Telephone speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First Predictions of CV syllable scores
  - First publish AI French and Steinberg 1947
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955
- Context:
  - G.A. Miller 1951 *Language and communication*
  - G.A. Miller 1962 5-word Grammar ≡ 4 dB of SNR
  - Boothroyd JASA 1968; Boothroyd & Nittrouer 1988
  - Bronkhorst et al. JASA 1993

2. Historical Speech Studies

- Lord Rayleigh 1908 and George Campbell 1910
  - Telephone speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First Predictions of CV syllable scores
  - First publish AI French and Steinberg 1947
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955
- Context:
  - G.A. Miller 1951 *Language and communication*
  - G.A. Miller 1962 5-word Grammar $\equiv$ 4 dB of SNR
  - Boothroyd JASA 1968; Boothroyd & Nittrouer 1988
  - Bronkhorst et al. JASA 1993

2. Historical Speech Studies

- Lord Rayleigh 1908 and George Campbell 1910
  - Telephone speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First Predictions of CV syllable scores
  - First publish AI French and Steinberg 1947
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955
- Context:
  - G.A. Miller 1951 *Language and communication*
  - G.A. Miller 1962 5-word Grammar ≡ 4 dB of SNR
  - Boothroyd JASA 1968; Boothroyd & Nittrouer 1988
  - Bronkhorst et al. JASA 1993

2. Historical Speech Studies

- Lord Rayleigh 1908 and George Campbell 1910
  - Telephone speech studies
- Harvey Fletcher's 1921 Articulation Index AI
  - First Predictions of CV syllable scores
  - First publish AI French and Steinberg 1947
- Shannon The theory of Information 1948+
  - G.A. Miller, Heise and Lichten Entropy $\mathcal{H}$ 1951
  - G.A. Miller & Nicely CM $P_{h|s}(SNR)$ 1955
- Context:
  - G.A. Miller 1951 *Language and communication*
  - G.A. Miller 1962 5-word Grammar $\equiv$ 4 dB of SNR
  - Boothroyd JASA 1968; Boothroyd & Nittrouer 1988
  - Bronkhorst et al. JASA 1993

2. Consonant Feature vs. Cue Studies

- Feature & Cue identification 1950-1990
  - 1950-1980: Haskins Labs (Features)
  - 1970-1990: MIT, Bell Labs, IU (Features)
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise
- Cochlear processing 1920-2000
  - 1910-1950: Bell Labs
  - 1960-2010: MIT + Harvard HSBT
  - 1980-2011: NIH funded University research
- Hearing Impaired CV studies (Cues)
  - 2004-2011: CV Confusion matrices Allen @ UIUC

2. Consonant Feature vs. Cue Studies

- Feature & Cue identification 1950-1990
  - 1950-1980: Haskins Labs (Features)
  - 1970-1990: MIT, Bell Labs, IU (Features)
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise
- Cochlear processing 1920-2000
  - 1910-1950: Bell Labs
  - 1960-2010: MIT + Harvard HSBT
  - 1980-2011: NIH funded University research
- Hearing Impaired CV studies (Cues)
  - 2004-2011: CV Confusion matrices Allen @ UIUC

2. Consonant Feature vs. Cue Studies

- Feature & Cue identification 1950-1990
  - 1950-1980: Haskins Labs (Features)
  - 1970-1990: MIT, Bell Labs, IU (Features)
  - 1980-2011: ASR research at AT&T, IBM, BBN, CMU
  - 2003-2011: NH Confusions in noise
- Cochlear processing 1920-2000
  - 1910-1950: Bell Labs
  - 1960-2010: MIT + Harvard HSBT
  - 1980-2011: NIH funded University research
- Hearing Impaired CV studies (Cues)
  - 2004-2011: CV Confusion matrices Allen @ UIUC

Allen's UIUC Experiments: 2004-2011

Year	Experiment	Student &Allen	Details	Publications
2004	MN04(MN64)	Phatak, Lovitt	MNR	JASA
2005	MN16R	Phatak, Lovitt	MN55R	JASA
2005	HIMCL05	Yoon, Phatak	10 HI ears	JASA
2006	HINALR05	Yoon *et al.*	10 HI ears	JSLR (2011)
2006	Verification	Regnier	/ta/	JASA
2006	CV06-s/w	Phatak/Regnier	8C+9V SWN/WN	
2007	CV06	Pan	CV06	MS Thesis
2007	HL07	Li	Hi/Lo pass	JASA
2008	TR08	Li	Furui86	ASSP
2009	3DDS	Li	plosives	JASA: TLSP
2009	Verification	Cvengros	burst mods	Thesis
2009	Verification	Abhinauv	burst mods	JASA
2009	mn64 NZE	Singh	PA07	JASA
2010	HIMCL10-I,II,III	Trevino, Han	46 HI ears @MCL	JASA/Sem Hear.
2011	3DDS	Li	Fricatives	JASA
2011	HINAL11-IV	Han	17 HI ears w NALR	PhD Thesis (Ch. 3)
2014	CV06	Toscano	30 NH ears	JSLHR

Allen's UIUC Experiments: 2004-2011

Year	Experiment	Student &Allen	Details	Publications
2004	MN04(MN64)	Phatak, Lovitt	MNR	JASA
2005	MN16R	Phatak, Lovitt	MN55R	JASA
2005	HIMCL05	Yoon, Phatak	10 HI ears	JASA
2006	HINALR05	Yoon *et al.*	10 HI ears	JSLR (2011)
2006	Verification	Regnier	/ta/	JASA
2006	CV06-s/w	Phatak/Regnier	8C+9V SWN/WN	
2007	CV06	Pan	CV06	MS Thesis
2007	HL07	Li	Hi/Lo pass	JASA
2008	TR08	Li	Furui86	ASSP
2009	3DDS	Li	plosives	JASA: TLSP
2009	Verification	Cvengros	burst mods	Thesis
2009	Verification	Abhinauv	burst mods	JASA
2009	mn64 NZE	Singh	PA07	JASA
2010	HIMCL10-I,II,III	Trevino, Han	46 HI ears @MCL	JASA/Sem Hear.
2011	3DDS	Li	Fricatives	JASA
2011	HINAL11-IV	Han	17 HI ears w NALR	PhD Thesis (Ch. 3)
2014	CV06	Toscano	30 NH ears	JSLHR

Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins Σ4
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs . . .) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)

Methods to Identify acoustic features

- Identify the key features in individual CV tokens MN55
  - -6 Plosives: /p, t, k/ and /b, d, g/
  - -8 Fricatives: U: /f, θ, s, ʃ/, V: /v, ð, z, ʒ/, + N: /m,n /
  - -With 4 vowels /ɑ, æ, ɛ, ɪ/
  - ≈18 talkers and >20-30 listeners
  - Up to 20 trials per consonant per SNR
- Methods: Speech processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

Methods to Identify acoustic features

- Identify the key features in individual CV tokens MN55
  - -6 Plosives: /p, t, k/ and /b, d, g/
  - -8 Fricatives: U: /f, θ, s, ʃ/, V: /v, ð, z, ʒ/, + N: /m,n /
  - -With 4 vowels /ɑ, æ, ɛ, ɪ/
  - ≈18 talkers and >20-30 listeners
  - Up to 20 trials per consonant per SNR
- Methods: Speech processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

Methods to Identify acoustic features

- Identify the key features in individual CV tokens MN55
  - -6 Plosives: /p, t, k/ and /b, d, g/
  - -8 Fricatives: U: /f, θ, s, ʃ/, V: /v, ð, z, ʒ/, + N: /m,n /
  - -With 4 vowels /ɑ, æ, ɛ, ɪ/
  - $\approx$18 talkers and $>$20-30 listeners
  - Up to 20 trials per consonant per SNR
- Methods: Speech processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

Methods to Identify acoustic features

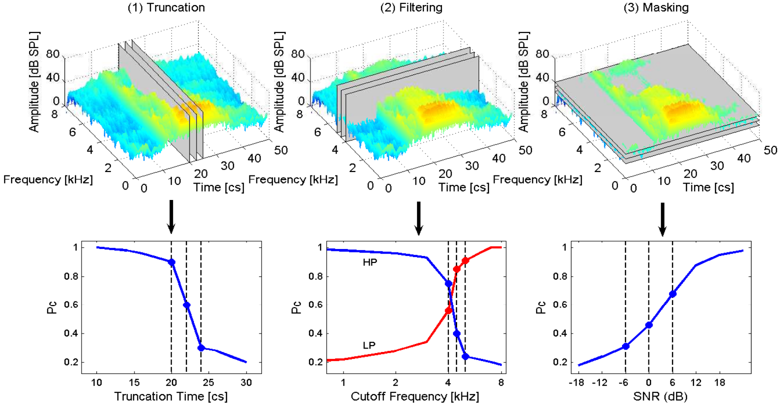
- Identify the key features in individual CV tokens MN55
  - -6 Plosives: /p, t, k/ and /b, d, g/
  - -8 Fricatives: U: /f, θ, s, ʃ/, V: /v, ð, z, ʒ/, + N: /m,n /
  - -With 4 vowels /ɑ, æ, ɛ, ɪ/
  - $\approx$18 talkers and $>$20-30 listeners
  - Up to 20 trials per consonant per SNR
- Methods: Speech processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

Methods to Identify acoustic features

- Identify the key features in individual CV tokens MN55
  - -6 Plosives: /p, t, k/ and /b, d, g/
  - -8 Fricatives: U: /f, θ, s, ʃ/, V: /v, ð, z, ʒ/, + N: /m,n /
  - -With 4 vowels /ɑ, æ, ɛ, ɪ/
  - ≈18 talkers and >20-30 listeners
  - Up to 20 trials per consonant per SNR
- Methods: Speech processing
  - AIgram Régnier & Allen 2008; Li & Allen 2009,10,11
- Method: 3^d Deep-Search (3DDS) via *truncations:*
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)

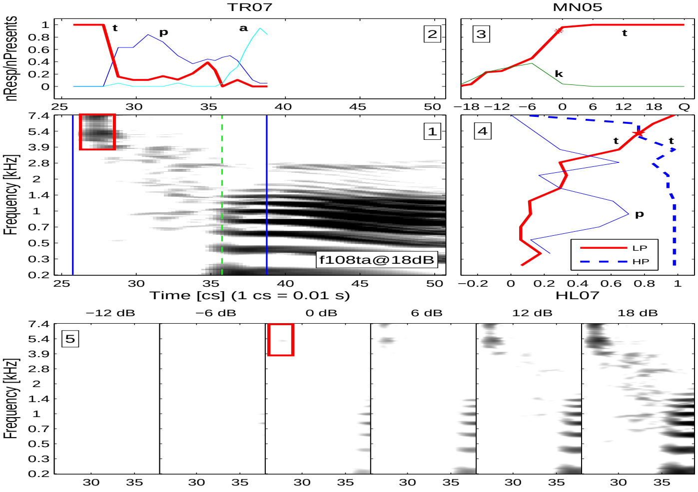
Methods: 3^d Deep Search (3DDS)

- 3^d Deep-Search (3^d-DS) via truncation:
  - Time truncation Furui 1986
  - Intensity truncation (i.e., masking)
  - Frequency truncation (High/Low-pass filtering)



3DDS Method /ta/

- Algram: Truncation in Time, Intensity and Frequency

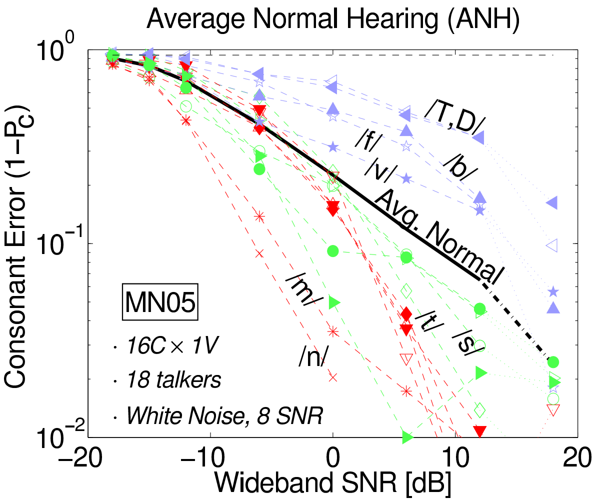


Outline

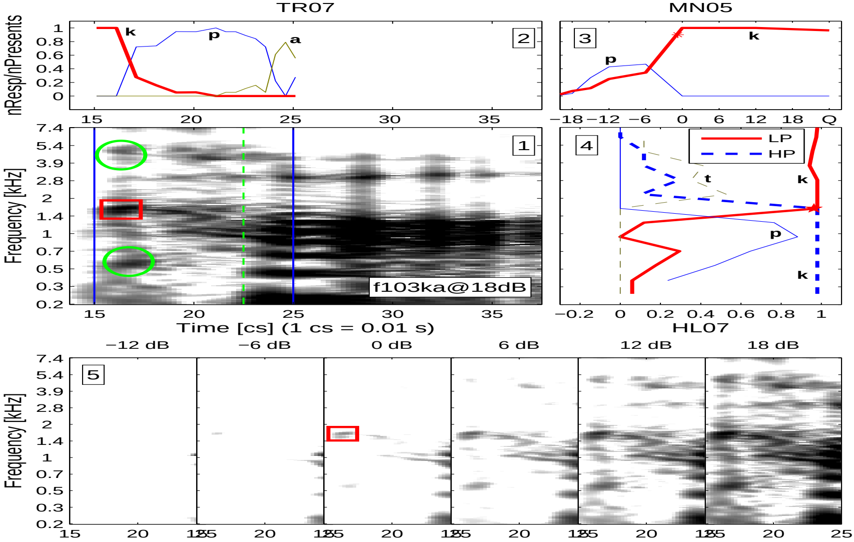
- Intro + Objectives — 2 mins
  - Research objectives
- Historical overview — 2 mins Σ4
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies (>1950)
- Methods — 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears — 5 mins Σ13
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his "missing bands"

4. Results NH ears [Phatak et al., 2008]

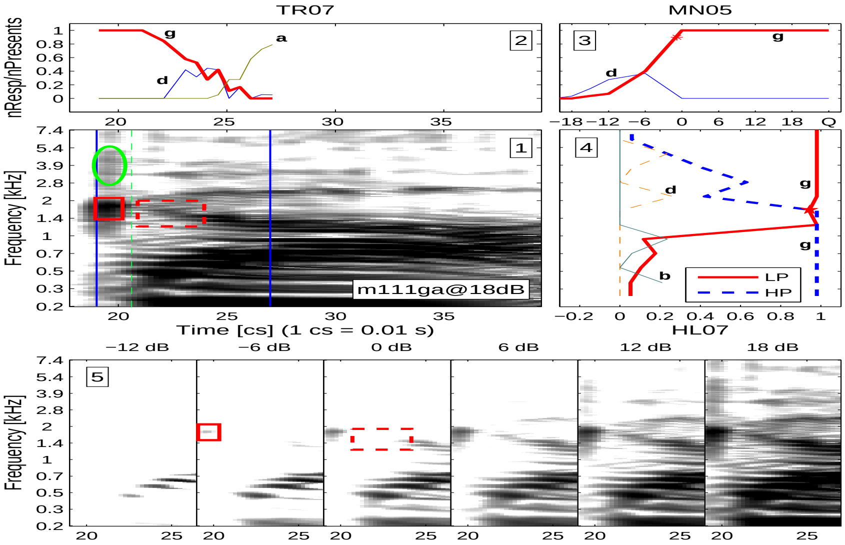
- Averaging obscures the multimodal error distribution



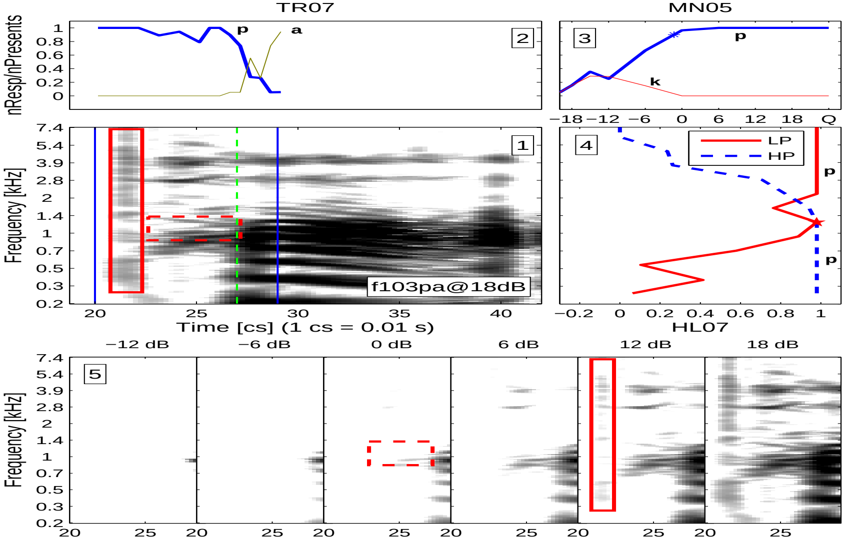
3DDS Method /ka/



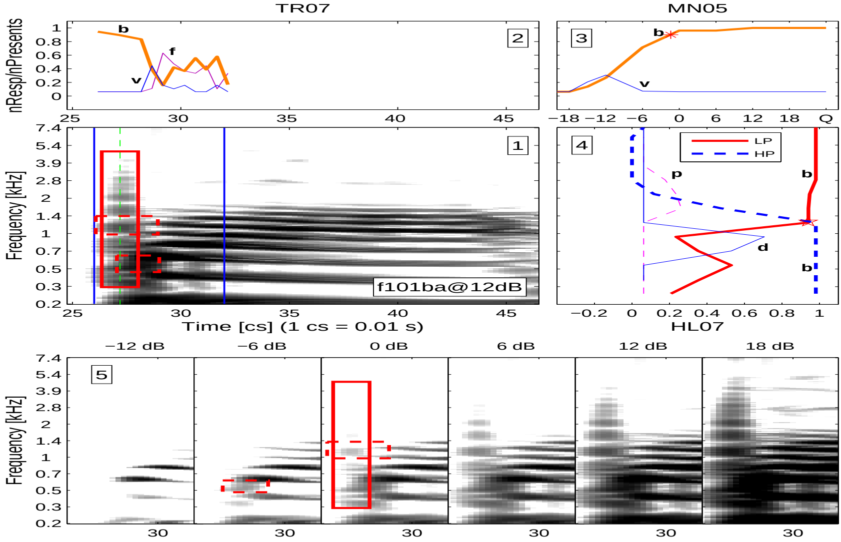
3DDS Method /ga/



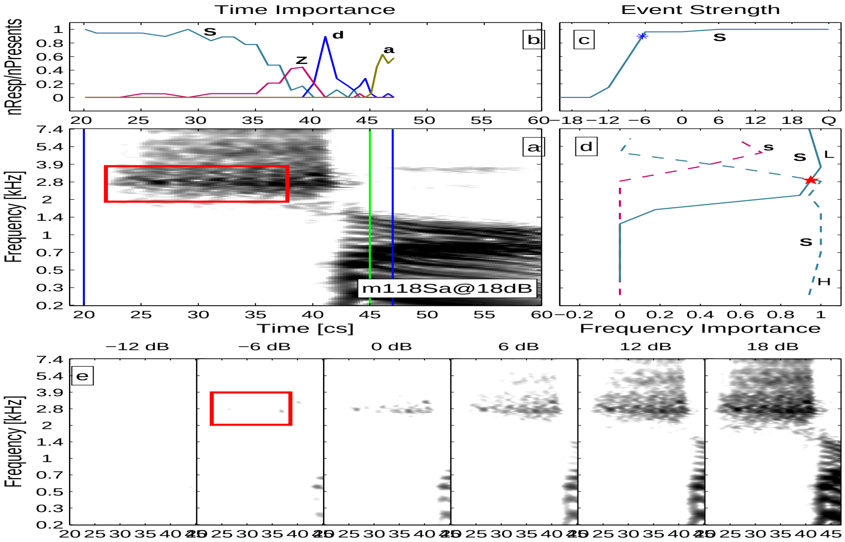
3DDS Method /pa/



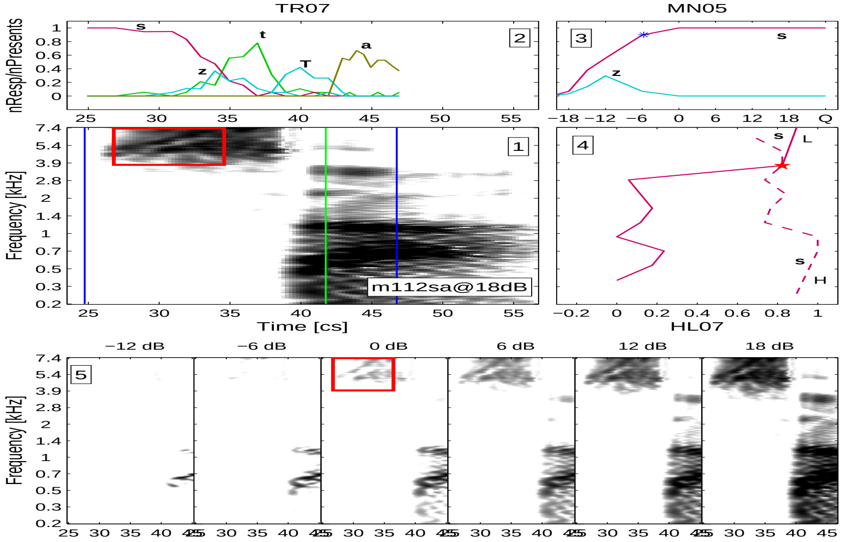
3DDS Method /ba/



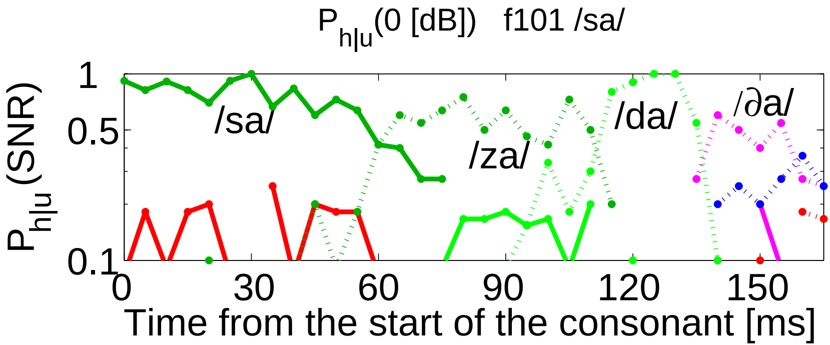
3DDS Method /ʃɑ/



3DDS Method /sa/



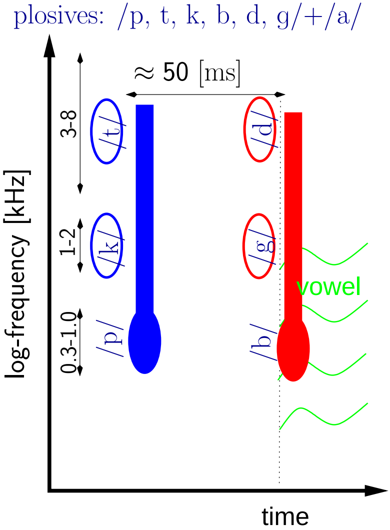
Truncation of f101 /sɑ/

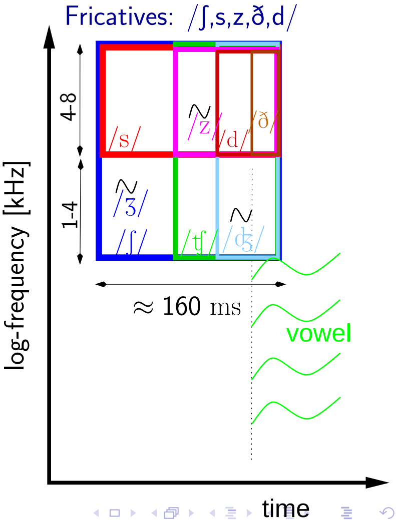


- NH responses to onset truncation /sɑ/
- Morphing from /sɑ/ → /zɑ/ → /dɑ/ → /ðɑ/
- Duration and low-frequency edge define fricative cues

Summary of Consonant structure

- Time-frequency structure of plosives and fricatives



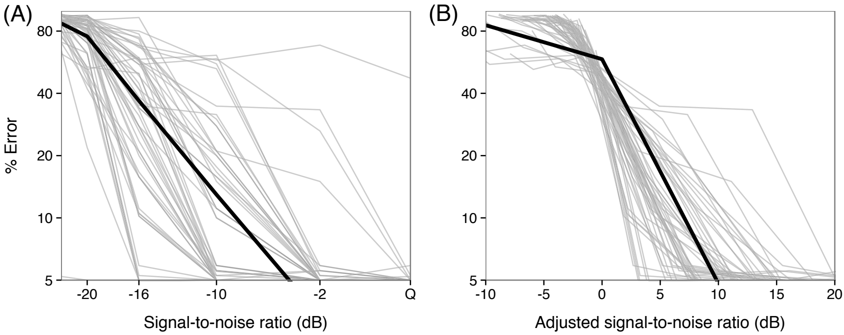


Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins Σ4
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies (>1950)
- Methods 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins Σ13
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his "missing bands"

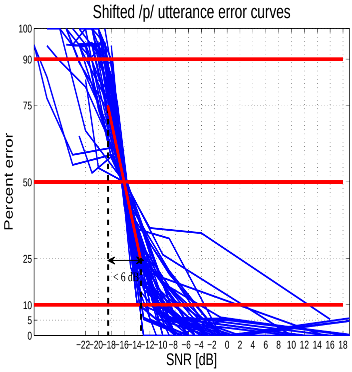
Error plots for all 56 /pV/ tokens

LEFT: $P_e(snr))$ RIGHT: $P_e(snr - snr_{50})$

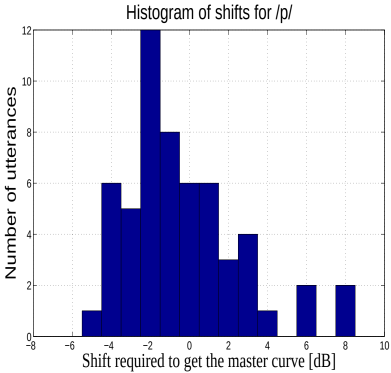


Properties of $P_e(SNR)$ for /p/

- Error vs. SNR shifted to 50% threshold SNR_{50} (LEFT)
- Histogram of SNR_{50} error thresholds (RIGHT)
  - Sharp transition ⇒ Binary Plosive identification!



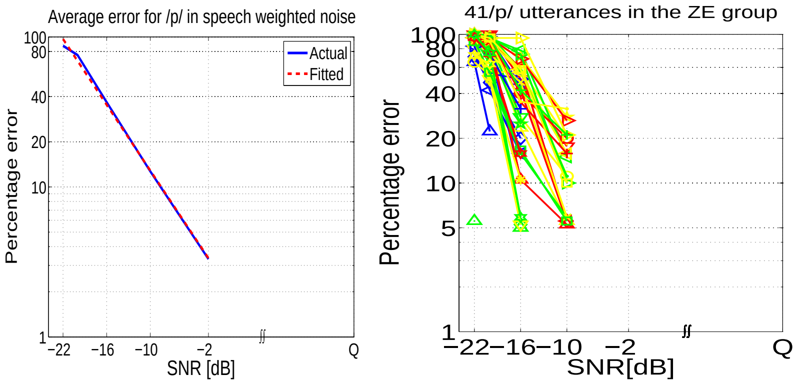
(a) $P_e(SNR - SNR^*_{50})$



(b) Distribution of SNR^*_{50}

Error plots for all 56 /pV/ tokens

- "Fitted" error: $log(P_e) = \alpha + \beta \cdot SNR$ (LEFT)
- Token errors $P_e(SNR)$ for 41 ZE /pV/ tokens (RIGHT)



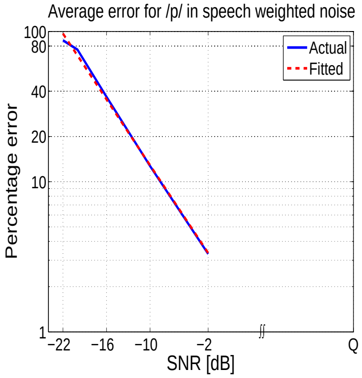
(a) /p/ log-error vs. SNR + model fit

(b) Zero-error group

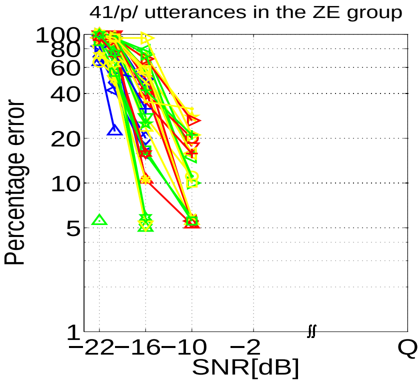
- This explains Fletcher's linear-log-error formula

Error plots for all 56 /pV/ tokens

- "Fitted" error: $log(P_e) = \alpha + \beta \cdot SNR$ (LEFT)
- Token errors $P_e(SNR)$ for 41 ZE /pV/ tokens (RIGHT)



(a) /p/ log-error vs. SNR + model fit



(b) Zero-error group

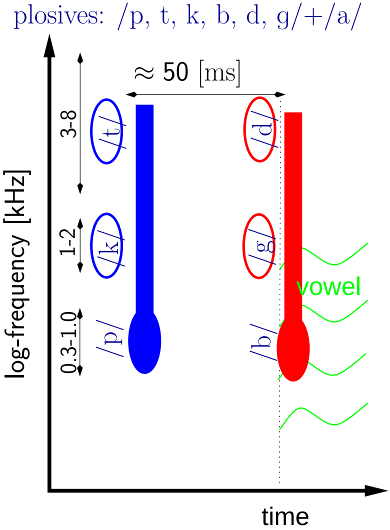
- This explains Fletcher's linear-log-error formula

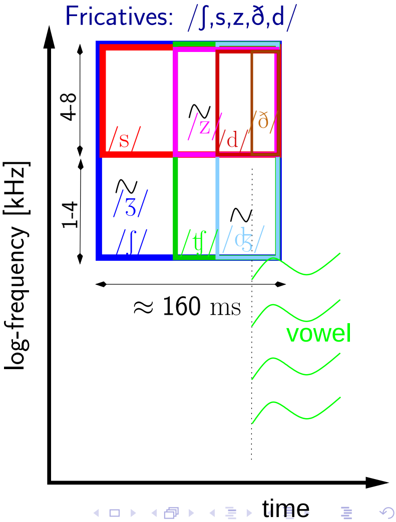
Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins Σ4
  - AG Bell (1860), Rayleigh (1910) to Shannon (1948)
  - Speech-feature studies (>1950)
- Methods 4 mins Σ8
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins Σ13
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter's "missing bands" ?

Summary of Consonant structure

- Time-frequency structure of plosives and fricatives





Outline

- Intro + Objectives 2 mins
  - Research objectives
- Historical overview 2 mins $\Sigma 4$
  - AG Bell (1860), Rayleigh (1908), Campbell (1910)
  - Fletcher (1921) to Shannon (1948)
  - Speech-feature studies (Haskins Labs . . .) (>1950)
  - G.A. Miller 1951+
- Methods 4 mins $\Sigma 8$
  - Psychophysics of speech/AIgram/3DDS
  - Under the hood of the Articulation Index?
  - Signal processing (STFT+AIgram)
- Results with NH ears 5 mins $\Sigma 13$
  - Multi-modal scores
  - Cues; Confusions; Primes and Morphs;
  - Binary nature of consonant percpetion
  - Kryter and his “missing bands”
- Summary + Conclusions 2 mins $\Sigma 15$

Conclusion I

- The average phone score is quantified by the articulation index
  - $P_c(SNR) = 1 - .02^{AI}$ is very accurate, but ...
  - is almost meaningless due to its large variance
- This AI variance depends on 3 main factors:
  - Across-Consonant error (20 dB spread)
  - Within-consonant: Utterance dependent thresholds
  - Confusions within each token below its SNR_{90}

Conclusion I

- The average phone score is quantified by the articulation index
  - $P_c(SNR) = 1 - .02^{AI}$ is very accurate, but ...
  - is almost meaningless due to its large variance
- This AI variance depends on 3 main factors:
  - Across-Consonant error (20 dB spread)
  - Within-consonant: Utterance dependent thresholds
  - Confusions within each token below its SNR_{90}

Conclusion I

- The average phone score is quantified by the articulation index
  - $P_c(SNR) = 1 - .02^{AI}$ is very accurate, but ...
  - is almost meaningless due to its large variance
- This AI variance depends on 3 main factors:
  - Across-Consonant error (20 dB spread)
  - Within-consonant: Utterance dependent thresholds
  - Confusions within each token below its SNR_{90}

Conclusion II

- We have identified specific consonant cues for 14 Cons !/m,n/

New tools:

1. AI-gram (cochlea-gram) to visualize speech features
2. 3DDS (truncate: time, freq, intensity) to isolated cues:
   Plosives /p, t, k/, /b, d, g/
   + Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/)
   + vowels /o, e, ɪ/
   - 18 different talkers, 96 CVs, $N_{trials} \geq 20$
3. To discriminate consonants in noise, NH listeners use
   - Plosives: *Burst + timing to Voicing*
   - Fricatives: *Low-frequency edge + duration*

Conclusion II

- We have identified specific consonant cues for 14 Cons !/m,n/ New tools:

1. AI-gram (cochlea-gram) to visualize speech features
2. 3DDS (truncate: time, freq, intensity) to isolated cues: Plosives /p, t, k/, /b, d, g/
   + Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/)
   + vowels /o, e, ɪ/
   - 18 different talkers, 96 CVs, $N_{trials} \geq 20$
3. To discriminate consonants in noise, NH listeners use
   - Plosives: *Burst + timing to Voicing*
   - Fricatives: *Low-frequency edge + duration*

Conclusion II

- We have identified specific consonant cues for 14 Cons !/m,n/
  New tools:

1. AI-gram (cochlea-gram) to visualize speech features
2. 3DDS (truncate: time, freq, intensity) to isolated cues:
   Plosives /p, t, k/, /b, d, g/
   + Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/)
   + vowels /o, e, ı/
   - 18 different talkers, 96 CVs, $N_{trials} \geq 20$
3. To discriminate consonants in noise, NH listeners use
   - Plosives: *Burst + timing to Voicing*
   - Fricatives: *Low-frequency edge + duration*

Conclusion II

- We have identified specific consonant cues for 14 Cons !/m,n/

New tools:

1. AI-gram (cochlea-gram) to visualize speech features
2. 3DDS (truncate: time, freq, intensity) to isolated cues:
   Plosives /p, t, k/, /b, d, g/
   \+ Fricatives /θ, ʃ, ʧ, s, h, f/, /z, ʒ, v, ð/)
   \+ vowels /o, e, ı/
   - 18 different talkers, 96 CVs, $N_{trials} \geq 20$
3. To discriminate consonants in noise, NH listeners use
   - Plosives: *Burst + timing to Voicing*
   - Fricatives: *Low-frequency edge + duration*

Conclusion III

We have demonstrated:

- Plosive cues are binary
  - The cue threshold is abrupt (i.e., 6 dB)
  - Thus consonants are NOT redundant
- How the AI works:
  - Burst, frequency-edge, timing & SNR_{50} distributions
  - $P_e(\mathrm{SNR}) = e_{\min}^{SNR}$ due to SNR_{50} distribution

Conclusion III

We have demonstrated:

- Plosive cues are binary
  - The cue threshold is abrupt (i.e., 6 dB)
  - Thus consonants are NOT redundant
- How the AI works:
  - Burst, frequency-edge, timing & SNR_{50} distributions
  - $P_e(SNR) = e_{min}^{SNR}$ due to SNR_{50} distribution

Conclusion III

We have demonstrated:

- Plosive cues are binary
  - The cue threshold is abrupt (i.e., 6 dB)
  - Thus consonants are NOT redundant
- How the AI works:
  - Burst, frequency-edge, timing & SNR_{50} distributions
  - $P_e(\mathrm{SNR}) = e_{\min}^{SNR}$ due to SNR_{50} distribution

Question your basic assumptions

Thank you for your attention

http://auditorymodels.org/
http://auditorymodels@Allen'sPublications

Bibliography

A. Boothroyd. Statistical theory of the speech discrimination score. *J. Acoust. Soc. Am.*, 43(2):362–367, 1968.

A. Boothroyd and S. Nittrouer. Mathematical treatment of context effects in phoneme and word recognition. *J. Acoust. Soc. Am.*, 84(1):101–114, 1988.

A. W. Bronkhorst, A. J. Bosman, and G. F. Smoorenburg. A model for context effects in speech recognition. *J. Acoust. Soc. Am.*, 93(1): 499–509, 1993.

G. A. Campbell. Telephonic intelligibility. *Phil. Mag.*, 19(6):152–9, 1910.

N. R. French and J. C. Steinberg. Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.*, 19:90–119, 1947.

S. Phatak, Andrew Lovitt, and Jont B. Allen. Consonant confusions in white noise. *J. Acoust. Soc. Am.*, 124(2):1220–33, 2008.

Lord Rayleigh. Acoustical notes – viii. *Philosophical Magazine*, 16(6): 235–246, 1908.

Conclusion II

- The Audiogram and Average Consonant errors have low correlation
- Averaging destroys important HI-subject information
  - We call this the 3 deadly SINs of averaging
    - Across *consonants:* SIN_c
    - Across *tokens:* SIN_u
    - Across *confusion:* SIN_f
- Utterances of a given consonant can have different
  - Errors
  - Confusions